

Relatório de Análise de Dados

Este relatório apresenta a análise de dados e técnicas de machine learning aplicadas aos dados fornecidos.

Análise de Componentes Principais (PCA)

A PCA (Análise de Componentes Principais) é uma técnica de redução de dimensionalidade que transforma os dados em um novo sistema de coordenadas. As novas coordenadas, ou componentes principais, são combinações lineares das variáveis originais.

Variáveis Consideradas para PCA:

Todas as variáveis numéricas presentes no dataset foram usadas para a análise PCA. Isso inclui variáveis como idade, número de infecções, e número de vacinas.

PCA:

Resultados da Regressão Linear

Métricas de Regressão Linear:

- Mean Squared Error (MSE):** {mse:.2f}
- R-squared (R^2):** {r2:.2f}

A regressão linear foi utilizada para prever a variável dependente com base nas variáveis independentes. O gráfico acima mostra a relação entre os valores reais e previstos pela regressão linear.

Regressão Linear:

Distribuição das Variáveis

O histograma mostra a distribuição dos dados, ajudando a entender a frequência das diferentes faixas de valores.

Histograma:

O boxplot ajuda a identificar outliers e a dispersão dos dados.

Boxplot:

O gráfico de barras da média das variáveis pode mostrar quais variáveis têm maiores médias, indicando tendências gerais.

Gráfico de Barras:

Conclusões

- Os componentes principais revelam as direções de maior variância nos dados, ajudando a identificar padrões importantes.
- A regressão linear fornece uma visão de como as variáveis independentes afetam a variável dependente.
- O histograma mostra a distribuição dos dados, ajudando a entender a frequência das diferentes faixas de valores.
- O boxplot ajuda a identificar outliers e a dispersão dos dados.
- O gráfico de barras da média das variáveis pode mostrar quais variáveis têm maiores médias, indicando tendências gerais.